# Sunil B

Email: banasunilkumar60@gmail.com

## **Professional Summary**

* Experienced **Data Engineer**with over **8.5 years of experience** with a proven track record in **healthcare data integration**, working with standards like **HL7, FHIR, CDA, DICOM**, and **ICD-10/ICD-11**.
* Skilled in **ETL pipeline design and optimization** using tools such as **Talend**, **Informatica**, **AWS Glue**, and **Azure Data Factory** for robust data movement and transformation.
* Expertise in handling **ANSI X12 EDI transactions** (837, 834, 835), ensuring **HIPAA** and **5010 X12 compliance** in medical claims and eligibility workflows.
* Proficient in developing and tuning **complex SQL queries**, **stored procedures**, and **data transformations** to ensure data integrity and performance.
* Strong hands-on experience in **cloud-based architectures** using **AWS (S3, Lambda, Redshift)** and **Azure (Blob Storage, SQL Database, Synapse Analytics)**.
* Delivered **ICD-10 to ICD-11 migration** projects, leading architecture, data mapping, and validation, improving **coding accuracy by 15%**.
* Built interactive **Power BI** and **Tableau dashboards** on Azure to support real-time analytics in healthcare environments.
* Implemented **data validation and cleansing frameworks**, increasing **data quality by up to 20%** across clinical and reporting systems.
* Adept at **Agile project management**, collaborating cross-functionally using **Jira** and **Trello** to deliver projects with **20–30% increased efficiency**.
* Led **User Acceptance Testing (UAT)** and compliance validation for data systems, ensuring conformance with **Medicaid** and **HIPAA** guidelines.
* Architected **scalable, high-performance data solutions**, enabling efficient processing of **large-scale healthcare datasets** in real-time.
* Experienced in **DevOps and CI/CD** practices using **Jenkins**, **Docker**, and **Kubernetes**, ensuring reliable deployment and version control.
* Developed and maintained **Medicaid data reporting pipelines**, cutting **processing times by 30%** and improving system automation.
* Contributed to **audit readiness** and **secure data workflows**, aligning with strict **healthcare compliance and security standards**.
* Successfully led **end-to-end SDLC** for multiple data engineering initiatives, from **requirements gathering** to **deployment and support**, ensuring delivery of high-impact, **data-driven healthcare solutions**.
* Demonstrated ability to **bridge the gap between technical and business teams**, translating complex data engineering concepts into actionable insights that **drive clinical and operational excellence**.
* Championed **real-time healthcare analytics** by integrating streaming data into reporting systems, empowering faster and more informed **clinical decision-making**.
* Consistently delivered **performance-optimized solutions** by identifying bottlenecks in **ETL workflows** and applying **query tuning** strategies, resulting in enhanced system responsiveness and reduced latency.
* Certified **AWS Data Engineer – Associate**, demonstrating validated expertise in **cloud-based data engineering solutions**.

## **Skills**

* **Programming**: Python, Java, SQL, No SQL
* **Data Integration & ETL:** Talend, Informatica, Apache Spark, AWS Glue, Azure Data Factory, SQL, Python, ETL Automation, Data Transformation, Data Pipelines
* **Cloud Platforms:** AWS S3, AWS Lambda, AWS Redshift, Azure Blob Storage, Azure SQL Database, Azure Synapse Analytics, Azure Data Factory, Cloud Data Solutions
* **Database Management:** SQL Server, MySQL, PostgreSQL, MongoDB, Azure SQL Database, Data Warehousing, Complex SQL Queries, Stored Procedures
* **Data Visualization:** Power BI (Azure), Tableau, Real-Time Dashboards, Reporting
* **Dev Ops &CI/CD:** Jenkins, Docker, Kubernetes, Github
* **Data Quality & Validation:** Data Cleansing, Validation, Automated Quality Checks, EDI Data Validation (837, 834, 835)
* **Project Management:** Agile (Jira, Trello), UAT Coordination, Cross-functional Team Collaboration, Communication, Documentation, C2C Client Collaboration
* **Performance Optimization:** ETL Workflow Optimization, Data Processing Efficiency, Query Optimization, Storage Performance
* **Healthcare Standards & Compliance:** HL7, FHIR, CDA, DICOM, CCD, CCR, ICD-10, ICD-11, Medicaid Data Reporting, EDI (837, 834, 835), HIPAA Compliance
* **Compliance & Security:** Medicaid Compliance, HIPAA, Data Security, Audit Readiness

## **Work Experience**

### **Senior Data Engineer**

**(C2C Contract) (Remote)-** State of NC,Raleigh, NC | Mar 2022 – Present

**Project Name: Healthcare Data Analysis & Integration**

* Designed and implemented data integration solutions for healthcare systems, utilizing HL7, FHIR, CDA, and DICOM standards to ensure smooth clinical data exchange.
* Developed systems to resolve patient identity mismatches, enhancing data accuracy and improving data quality across healthcare systems.
* Architected and deployed HIT frameworks (CCD, CCR) to improve interoperability, enabling faster clinical decision-making and reducing time-to-insight for operational teams.
* Led User Acceptance Testing (UAT) and validation to ensure compliance with healthcare data standards and operational requirements.
* Led cross-functional teams within an Agile framework, ensuring alignment between business objectives and technical requirements, which increased project efficiency by 20%.
* Coordinated with stakeholders to prioritize and define data engineering deliverables, ensuring timely and impactful outcomes throughout the SDLC.
* Managed the full lifecycle of **ETL pipeline development**, from data extraction to transformation and loading into centralized data repositories.
* Designed, built, and optimized **ETL pipelines** using Talend and Informatica to integrate data from multiple healthcare systems into centralized data warehouses.
* Automated and optimized **SQL** queries and **ETL workflows**, resulting in a 25% reduction in processing times and improved system performance.
* Developed complex **SQL** queries and stored procedures to ensure high-quality, consistent data for advanced analysis and reporting.
* Delivered scalable **data pipelines** capable of handling large healthcare datasets, ensuring real-time data availability for both clinical and operational teams.
* Ensured efficient data transformations, supporting **analytics**, reporting, and decision-making across the organization.
* Implemented integration technologies for seamless exchange of clinical data, driving operational efficiency and improving healthcare service delivery.
* Leveraged data engineering tools such as **Talend** and **Informatica** for data integration and transformation, supporting end-to-end data flow management.
* Designed and deployed scalable, high-performance data architectures to ensure smooth handling of large-scale healthcare data in real-time environments.

### **Senior Data Engineer**

**(C2C Contract) (Remote)-** Medical Group, Hartford, CT | Oct 2019 – Mar 2022

**Project Name: ICD-10 to ICD-11 Migration**

* Led the technical architecture and design for the migration from ICD-10 to ICD-11, ensuring seamless integration within clinical, billing, and reporting systems.
* Developed data validation processes to ensure the migrated ICD-10 data complied with ICD-11 standards, maintaining accuracy, consistency, and regulatory compliance.
* Created training materials and provided documentation to healthcare teams, improving the coding accuracy of ICD-11 by 15% across clinical operations.
* Coordinated with cross-functional teams, including software engineering and IT, to ensure smooth integration of the ICD-11 migration within existing data architectures, promoting operational continuity.
* Managed end-to-end SDLC of the migration process, from data extraction and transformation to final deployment of the new ICD-11 system.
* Delivered project milestones on time, ensuring adherence to deadlines and business requirements within an Agile framework.
* Designed and developed optimized **ETL processes** to transform and map ICD-10 codes to ICD-11, reducing manual processes by 30% and ensuring data accuracy.
* Built and maintained robust data pipelines to support ongoing data operations post-migration, ensuring continuous processing of ICD-11 data for reporting and analytics.
* Created complex **SQL** queries and scripts to verify data integrity and completeness throughout the ICD-10 to ICD-11 transition process, improving data validation efficiency.
* Utilized **Azure Data Factory** to orchestrate **ETL workflows**, ensuring efficient data migration and transformation from ICD-10 to ICD-11 with minimal errors.
* Leveraged **PowerBI** on **Azure** to develop real-time dashboards for stakeholders to track migration progress, focusing on key metrics such as data completeness and accuracy.
* Optimized database structures using **Azure SQL Database** and **Azure Synapse Analytics**, reducing query execution time by 20% and enhancing overall system performance during the migration.

### **Data Engineer**

**(C2C Contract) (On-Site)-** Biogen Pharma, Cambridge, MA | Jan 2018 – Oct 2019

**Project Name: Medicaid Data Reporting and Analytics**

* Led the development of Medicaid data reporting solutions, ensuring compliance with healthcare standards, including HIPAA and Medicaid guidelines.
* Collaborated with stakeholders to gather and define data requirements, translating business needs into technical specifications for Medicaid reporting systems.
* Created **Power BI** and **Tableaudashboards** to visualize key healthcare metrics, such as cost per patient and treatment success rates, enabling data-driven decision-making.
* Managed the full SDLC for Medicaid data reporting pipelines, from gathering requirements to implementation and ongoing maintenance.
* Coordinated with cross-functional teams to ensure data solutions met business needs and healthcare compliance standards.
* Documented the technical and business requirements, ensuring clear communication across teams and smooth project execution.
* Designed and optimized **ETL pipelines** for Medicaid claims data, improving report generation times by 30% and automating data processing tasks to increase operational efficiency by 25%.
* Implemented robust data validation and cleaning processes, improving data accuracy by 20% and ensuring reliable data for analysis and reporting.
* Developed and maintained **SQL** queries to extract, transform, and load Medicaid claims data into **AWS S3,** ensuring data accuracy and consistency across reports.
* Delivered scalable, automated data processing solutions, reducing manual intervention and improving the timeliness and efficiency of Medicaid reporting.
* Utilized **AWS S3** for data storage, optimizing partitioning strategies to enhance data retrieval and processing performance by 15%.
* Leveraged **AWS services** to ensure secure and compliant data storage and processing workflows, maintaining HIPAA and Medicaid compliance throughout the project.

### **Junior Data Engineer**

**(C2C Contract)(On-Site) -** Veeva Systems, Pleasanton, CA | Aug2016 – Jan 2018

**Project Name: EDI Data Integration and Reporting for Healthcare**

* Supported the integration of healthcare data through Electronic Data Interchange (EDI) transactions, specifically working with ANSI X12 transaction sets such as 837 (**claims**), 834 (**eligibility**), and 835 (payment/remittance) to improve the exchange of medical billing and claims data with external partners.
* Worked with healthcare stakeholders to gather and define EDI data requirements, ensuring compliance with industry standards like HIPAA and 5010 X12.
* Participated in the full **SDLC**, from initial requirements gathering to the implementation and maintenance of EDI-based data workflows.
* Collaborated with senior engineers to define project milestones, prioritize tasks, and deliver data engineering solutions in an Agile environment.
* Documented **EDI data flows**, transaction formats (837, 834, 835), and validation rules to ensure transparency and compliance with healthcare data standards.
* Designed and developed ETL processes for transforming and integrating EDI data, ensuring that transactions like **837 (claims), 834 (eligibility), and 835 (payment**) were processed efficiently into internal databases.
* Built automated data validation workflows to identify discrepancies in incoming EDI transactions, significantly improving data accuracy and reducing errors by 20%.
* Optimized **SQL** queries to extract, transform, and load EDI healthcare data, enabling seamless reporting and analytics on transaction performance and status.
* Assisted in creating reporting dashboards using **Azure-based tools** like **Power BI** to provide stakeholders with real-time insights on claims processing and data trends.
* Utilized Azure Data Factory to orchestrate and automate the flow of EDI files, enhancing the scalability and efficiency of data processing pipelines.
* Leveraged **Azure Blob Storage** for secure and compliant storage of EDI data, implementing partitioning strategies to optimize retrieval and performance of large healthcare datasets.

## **Certifications:**

* **AWS Certified Data Engineer - Associate**